A NON-LINEAR ESTIMATION PROBLEM

BU-204-M R. C. Mosteller August, 1965

Introduction

The estimation problem to be described here is an outgrowth of an attempt to characterize a parasite epidemic in sheep. It seems germane, therefore, to provide the relevant biological background so that the estimation problem may be more clearly understood.

The parasite is a blood-sucking worm which inhabits the stomach of the sheep and subsists, at least in part, on blood taken from the stomach walls. Eggs produced by the female worms are passed out of the sheep in feces and subsequently develop into larvae. The larvae may then be ingested by a grazing sheep and, after implanting themselves in the stomach walls, develop into adult worms which add to and sustain the sheep's worm population.

Although the propagation of the epidemic is dependent upon the worm-larvae cycle, the distribution of worm populations is apparently independent of the average number of larvae ingested by the hosts when this number is large (above one thousand per day say). There seems to be some control mechanism which inhibits larval development when the worm burden reaches the sheep's "capacity". (Some evidence which supports this notion is that certain breeding strains have uniformly low worm burdens while other strains exposed to the same epidemic have uniformly high worm burdens.) Similarly, the numbers of eggs passed by the sheep show little correlation with their worm burdens. Here again it is believed that some control factor (or concomitant variable) exists which regulates the number of eggs that can be produced by a given worm population.

Efforts to pin-point these two control factors have led to problems of the same essential nature. Therefore, for the sake of brevity, the following discussion will be concerned with the egg production problem only.

Biometrics Unit, Plant Breeding Department, Cornell University

Since the worms are blood-suckers, it seems reasonable to expect blood studies to provide a concomitant variable which would account for some of the large variation in egg production. The worms do not use all of the blood taken from the host, however; they are known to discard some of it. Thus, the observed small correlation between sheep blood loss and egg production is to be expected. Furthermore, volume losses of all blood elements, such as red blood cells and blood sugar, believed to be important in egg production show small correlations with eggs produced. But is this not also to be expected? If both red blood cells and blood sugar are important ingredients in egg production then they are probably needed in nearly exact proportions, e.g., one unit of red blood cells for every two units of blood sugar. When oxygen from the red cells is needed blood sugar would be a poor substitute. Thus, in a particular sheep, if blood sugar concentration is low and red cell concentration high, the parasites will take many more red cells from the host than are used in egg production, while, in some other sheep in which the relative concentrations are reversed the ratio of eggs produced to red cells lost may be quite high. It is evident, therefore, that no one of the blood elements will be sufficient to explain the variation in egg production, indeed all or nearly all relevant blood elements must be considered together.

One possible approach to this problem is to attempt to estimate a blood factor or "egg factor" which is highly correlated with egg production. The egg factor which I have in mind is the formula in terms of blood elements which is used to produce one unit of egg mass. At this point a rather worn analogy might prove helpful. If the worms produced water rather than eggs and fed on hydrogen and oxygen in various relative concentrations rather than blood, and if they used up one oxygen atom in producing one water molecule and used neither hydrogen nor oxygen for any other purpose, then the "water factor" sought would be H_2O_2. Note that the water factor is not H_2O, and input of one unit of H_2O, under the stated conditions, would not result in an output of one water molecule. The water factor input should be highly correlated with water output, while the inputs of the elements hydrogen and oxygen may show little correlation since some quantities of one or the other would be discarded.

The problem of finding an egg factor is complicated by the fact that the worms do use some of the blood elements to maintain themselves. Here again, the notion of a blood factor of nearly exact relative proportions, a "maintenance factor", might be the most fruitful approach to determining the worms' maintenance needs. After all, the worms' activities should be about the same from hour to hour, their maintenance needs should be about the same, and when they need a particular substance there is probably only one blood element from which they can get it. The egg factor is to contain the blood elements needed to produce the eggs as well as those elements used in the eggs, so that the maintenance factors for both male and female worms should be about the same. On this basis, it may be possible to separate the two factors. But this conjecture can be explored only after a solution is found for the basic estimation problem.

Consider a transducer having a variable n-dimension vector input, X, and a single output, M. Let the input at the i^{th} time be denoted by

$$X_i = (x_{1i}, x_{2i}, x_{3i}, \ldots, x_{ni}) ,$$

where x_{ki} is the magnitude of the k^{th} component of the input at i^{th} time, and let M_i represent the magnitude of the corresponding output at the i^{th} time. Further, let

$$Y = (y_1, y_2, y_3, \ldots, y_n)$$

be an unknown vector of constants. Then, if the output at the j^{th} time is M_j, the input may be represented as

$$X_j' = M_j Y' + Z_j' + \epsilon_j , \qquad (1)$$

where

$$Z = (z_1, z_2, \ldots, z_n)$$

is an unknown variable vector, and ϵ_j is an unknown error vector. Equation (1) says that, if the output has magnitude M_j, the input X_j must contain exactly, within the limits of error, M_j units of the vector Y, but will also contain an unknown residual vector Z_j. The residual vector will contain no Y units, thus, among the elements of Z, which are non-negative, there must be at least one zero element.

The problem, then, is to specify minimal conditions which must be imposed on the Z_i and ϵ_i in order that an estimate of Y may be obtained, when given n observations on the X_i and M_i. Or at this point, the problem might be reduced to that of finding any reasonable estimation scheme.